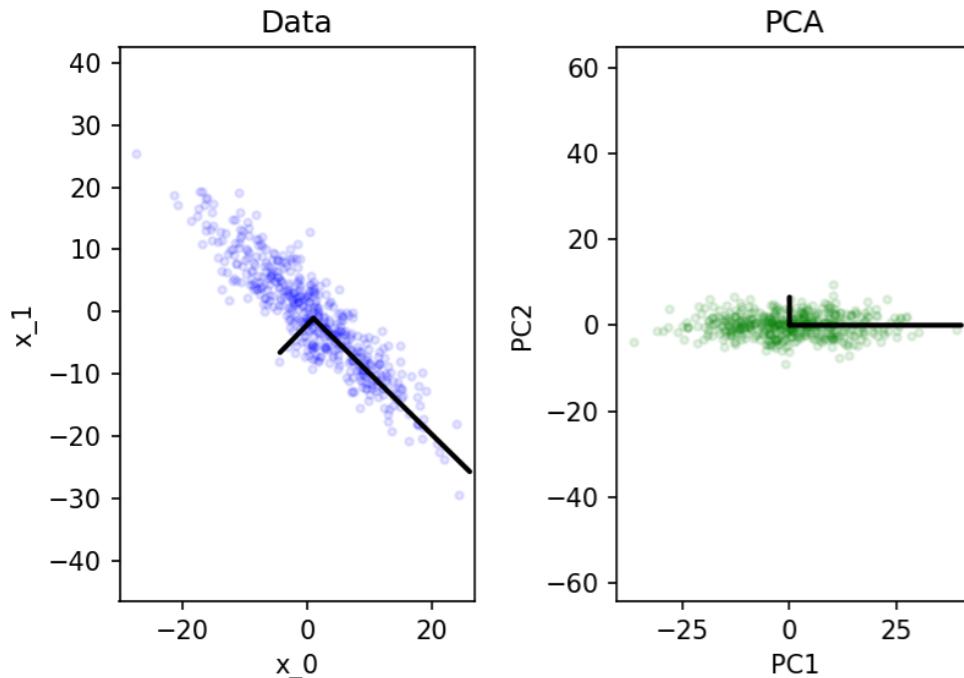
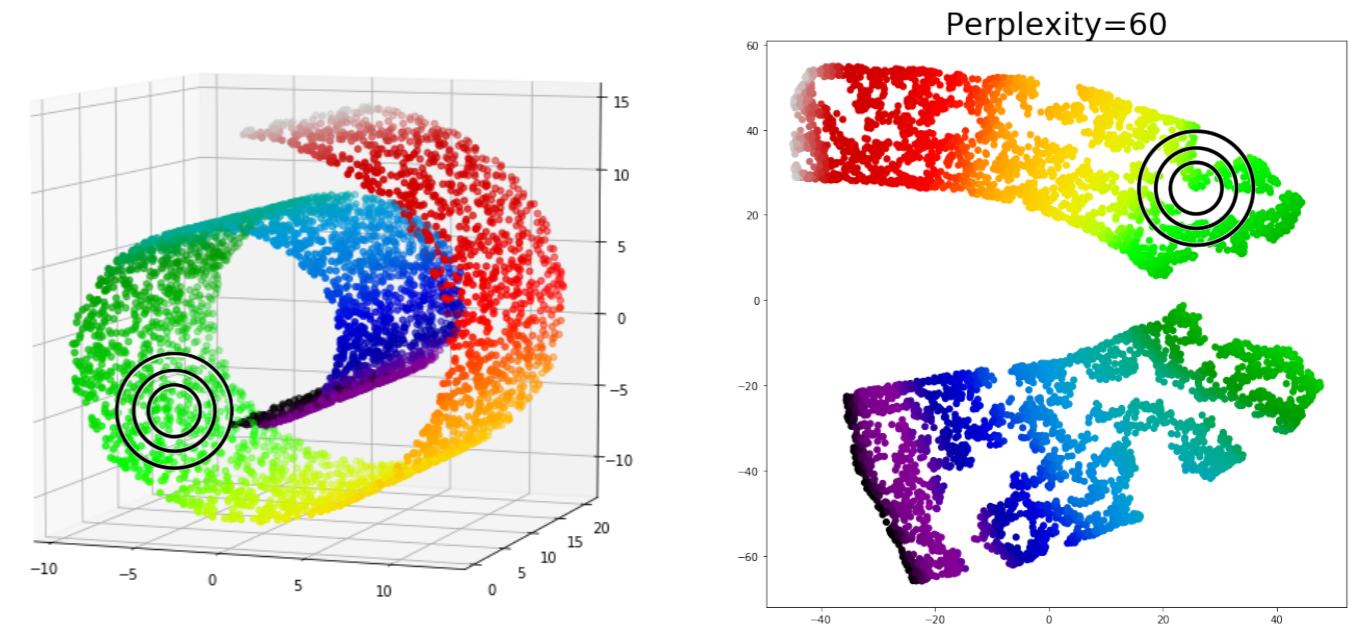


# Dimensionality reduction

## Principal Components Analysis



## t-Stochastic Neighbour Embedding



## Sparse Factor Analysis

$$Y = \Lambda X + \varepsilon$$

A diagram illustrating Sparse Factor Analysis. On the left is a matrix labeled  $Y$ . To its right is an equals sign. To the right of the equals sign are two matrices: one labeled  $\Lambda$  and one labeled  $X$ , separated by a plus sign. To the right of the plus sign is another equals sign. To the right of the second equals sign is a matrix labeled  $\varepsilon$ .

$$\begin{matrix} & \\ & \end{matrix} = \begin{matrix} & \\ & \end{matrix} + \begin{matrix} & \\ & \end{matrix} + \begin{matrix} & \\ & \end{matrix}$$

A diagram illustrating the sparse factor analysis equation using block matrices. It shows a large matrix on the left being decomposed into three smaller matrices on the right, each multiplied by a plus sign and followed by another plus sign.

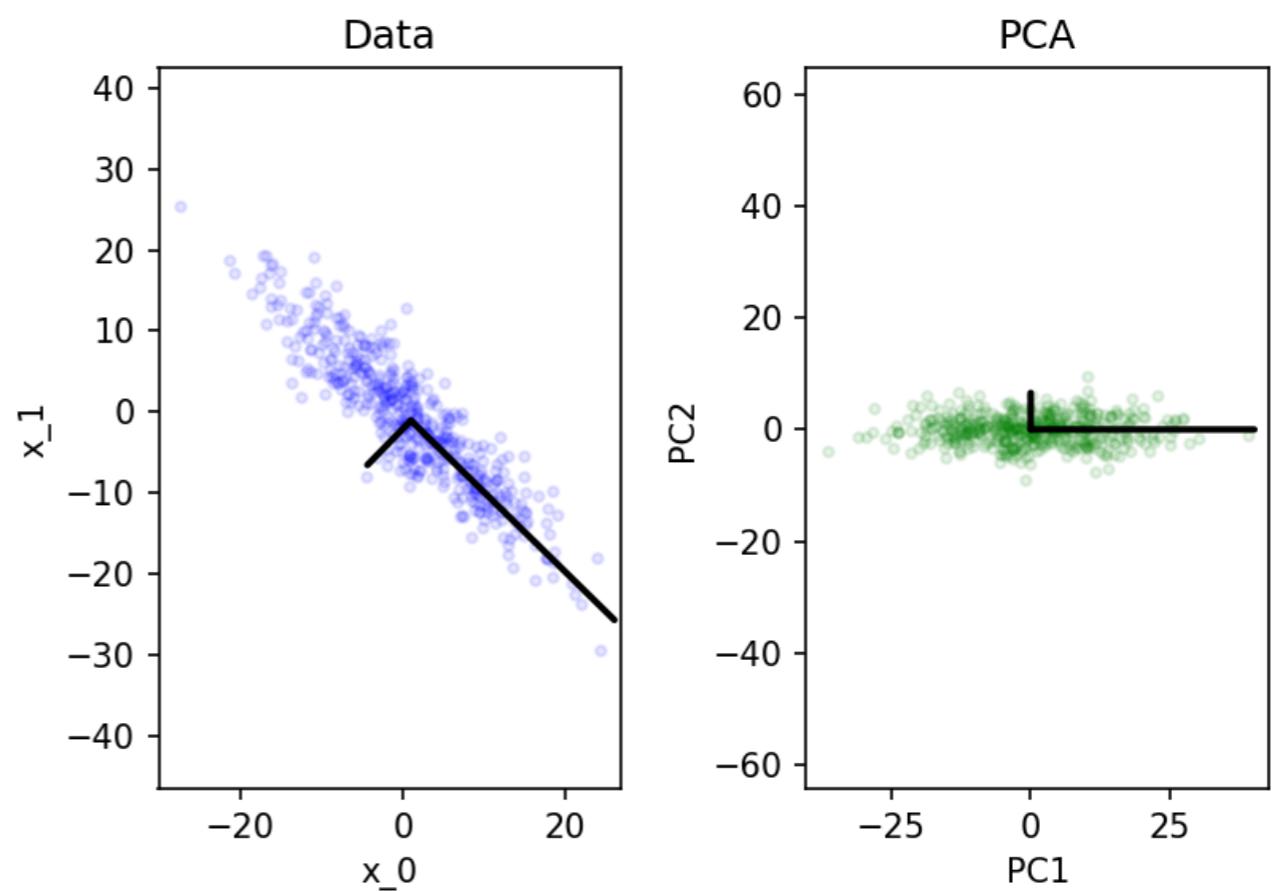
Kath Nicholls

# Aims of dimensionality reduction

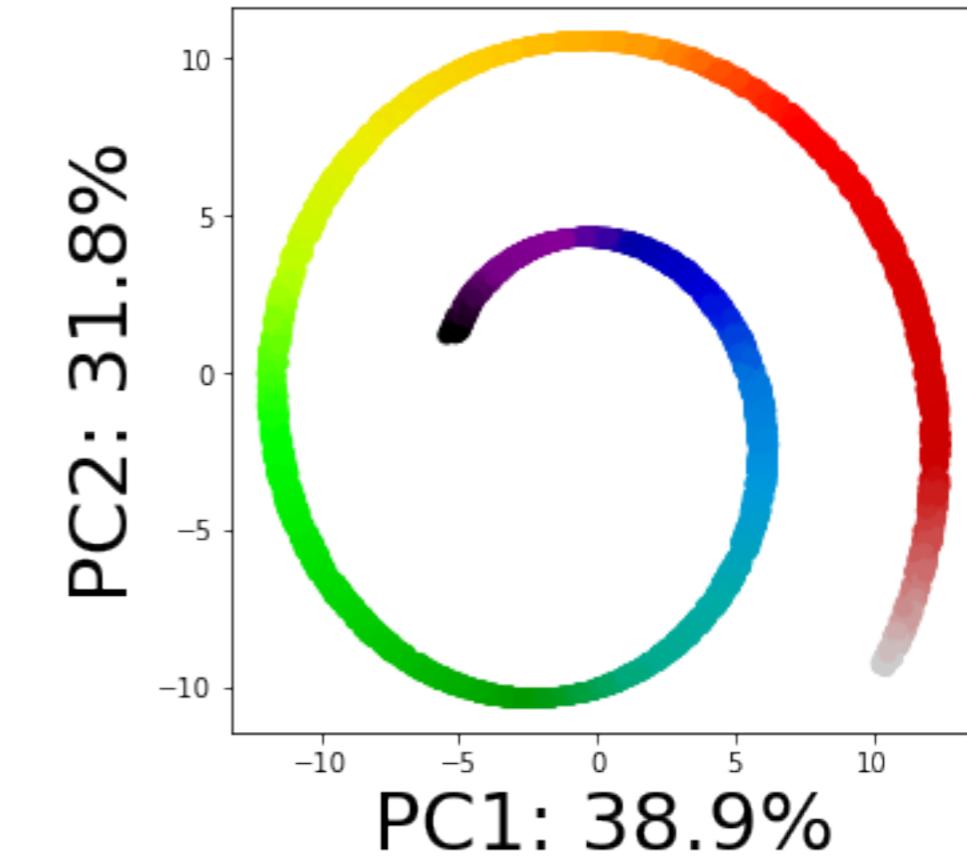
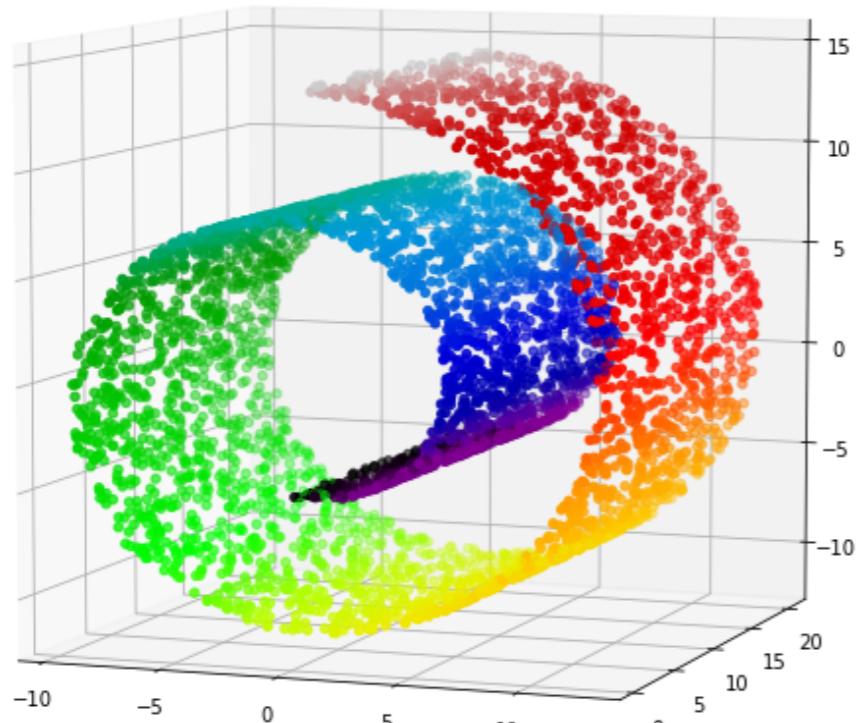
- Produce mapping to reduce dimensions without losing structure
- Visualise high-dimensional data
- Identify important variables
- Cluster data points (e.g. to check for a batch effect)

# Principal Components Analysis

- Directions of most variance
  - Linear
  - Orthogonal
- Explicit mapping
- Easy to change k
- Calculate using SVD or eigen decomposition



# Linearity of PCA



- PCA won't find non-linear directions of variance

# PCA via eigen decomposition

- $X$  has  $n$  rows (samples) and  $p$  columns (variables)
- $C = \frac{1}{n-1}X^T X$  is  $p \times p$  **covariance matrix**
- Eigen decomposition is  $C = VLV^T$

$$C = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_p \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ - & v_2^T & - \\ \vdots & & \\ - & v_p^T & - \end{bmatrix} \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

- $X \mapsto XV$  transforms to rotated  $p$ -dimensional space
- $X \mapsto XV_{1:k}$  transforms to *reduced*  $k$ -dimensional space
- $XV_j$  is called  $j$ th *principal component*

# PCA via singular value decomposition

- Singular value decomposition of  $X$  is

$$X = U D V^T$$
$$X = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_n^2 \end{bmatrix} \begin{bmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \lambda_n^2 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_p^T \end{bmatrix}$$

Eigenvalues of  $XX^T$

Eigenvalues of  $X^T X$

- Linked to eigen decomposition
- $XV_j = U_j D_j$  is jth principal component

# Reconstructing X

- Recall  $X = UDV^T$
- $V_{1:k}V_{1:k}^T \approx I$  (exact for k=p since V orthogonal)

$$X \approx X V_{1:k} V_{1:k}^T$$

- Link to factor analysis:

$$X = F\Lambda + \epsilon$$

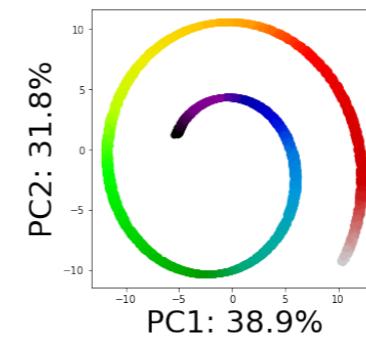
# Principal Components Analysis

- Directions of most variance

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

- Linear

- Orthogonal



- Explicit mapping

$$X \mapsto X V_{1:k}$$

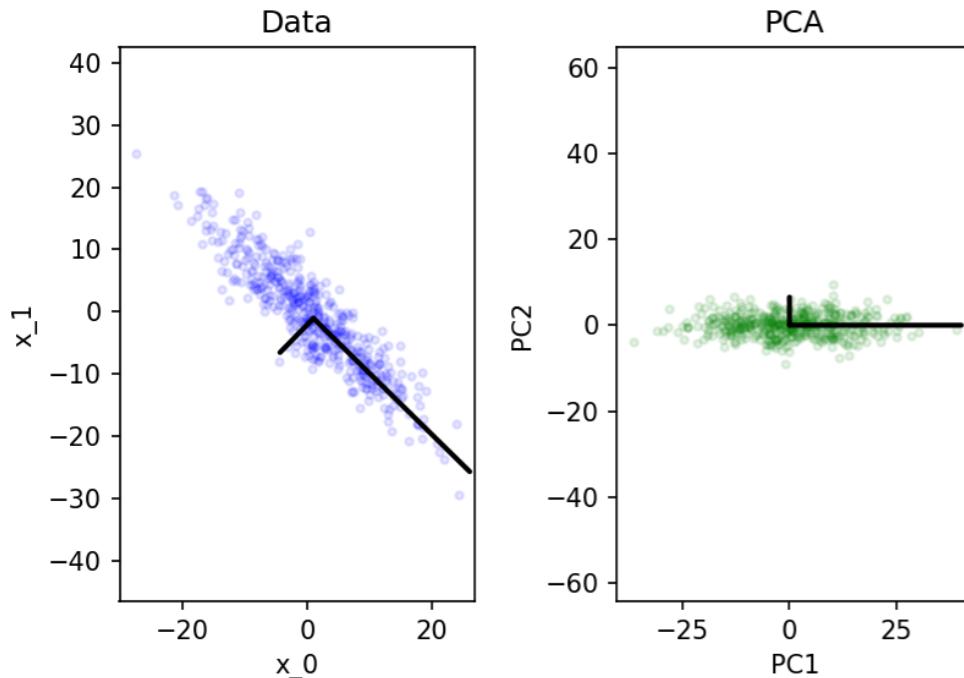
- Easy to change k

$$X \mapsto X V_{1:k}$$

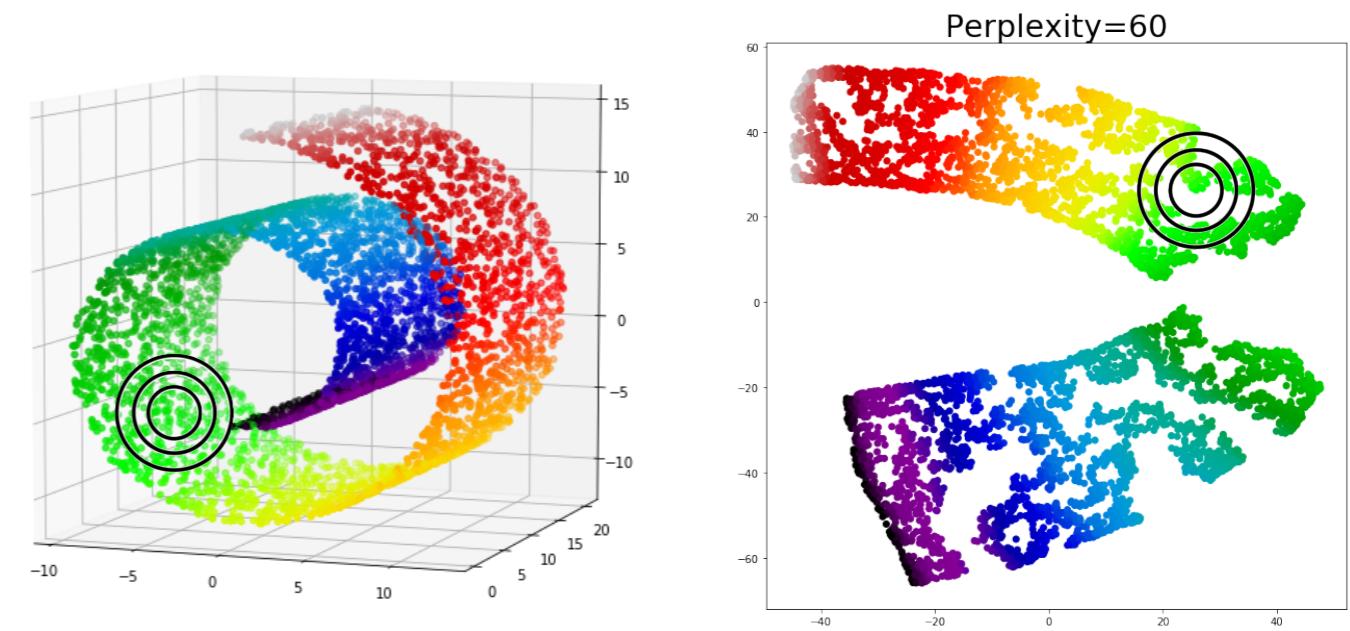
- Calculate using SVD or eigen decomposition

# Dimensionality reduction

## Principal Components Analysis



## t-Stochastic Neighbour Embedding



## Sparse Factor Analysis

$$Y = \Lambda X + \varepsilon$$

A diagram illustrating Sparse Factor Analysis. On the left is a matrix labeled  $Y$ . To its right is an equals sign. To the right of the equals sign are two matrices: one labeled  $\Lambda$  and one labeled  $X$ , separated by a plus sign. To the right of the plus sign is another equals sign. To the right of the second equals sign is a matrix labeled  $\varepsilon$ .

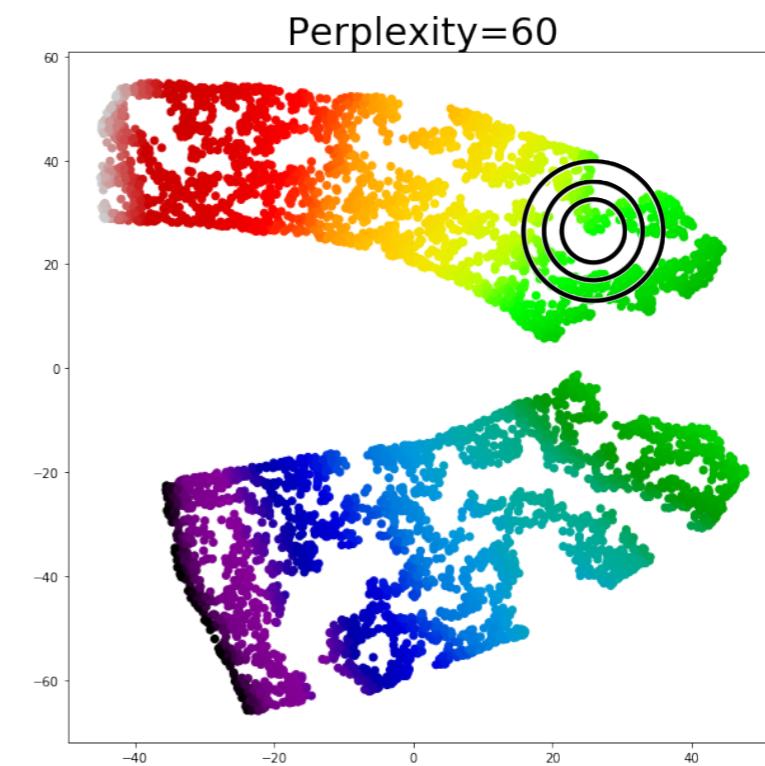
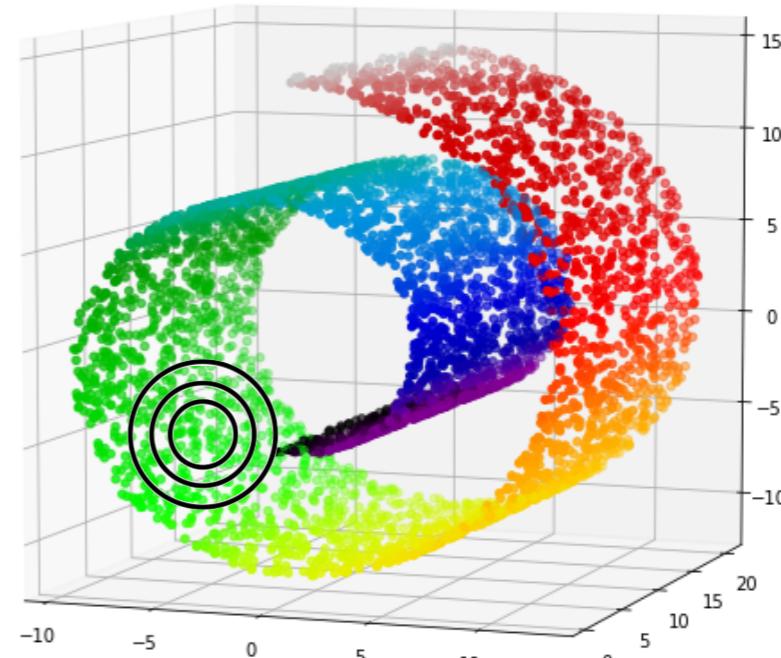
$$\begin{matrix} & \\ & \end{matrix} = \begin{matrix} & \\ & \end{matrix} + \begin{matrix} & \\ & \end{matrix} + \begin{matrix} & \\ & \end{matrix}$$

A diagram illustrating the sparse nature of the factor analysis model. It shows a large matrix divided into four quadrants. The top-left quadrant is filled with blue pixels, while the other three quadrants are mostly white. This is followed by a plus sign, then three smaller matrices: the top-right is mostly white with a few blue pixels, the bottom-left is mostly white with a few blue pixels, and the bottom-right is mostly white with a few blue pixels.

Kath Nicholls

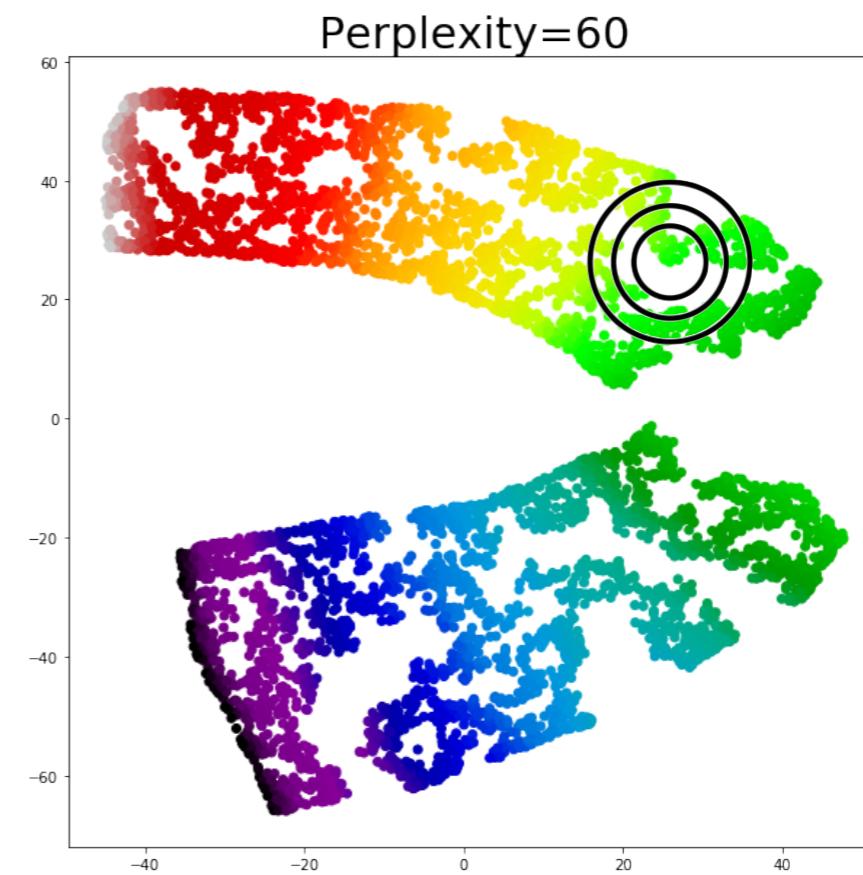
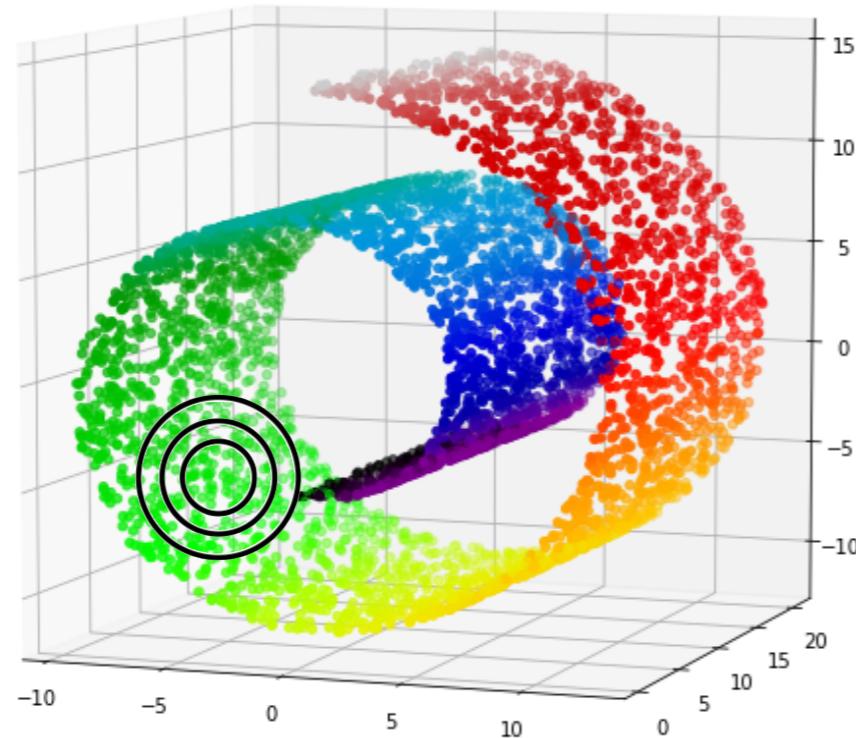
# t-distributed Stochastic Neighbour Embedding

- Preserves local similarity
- Stochastic
- Hyperparameters
  - Perplexity
- No explicit mapping



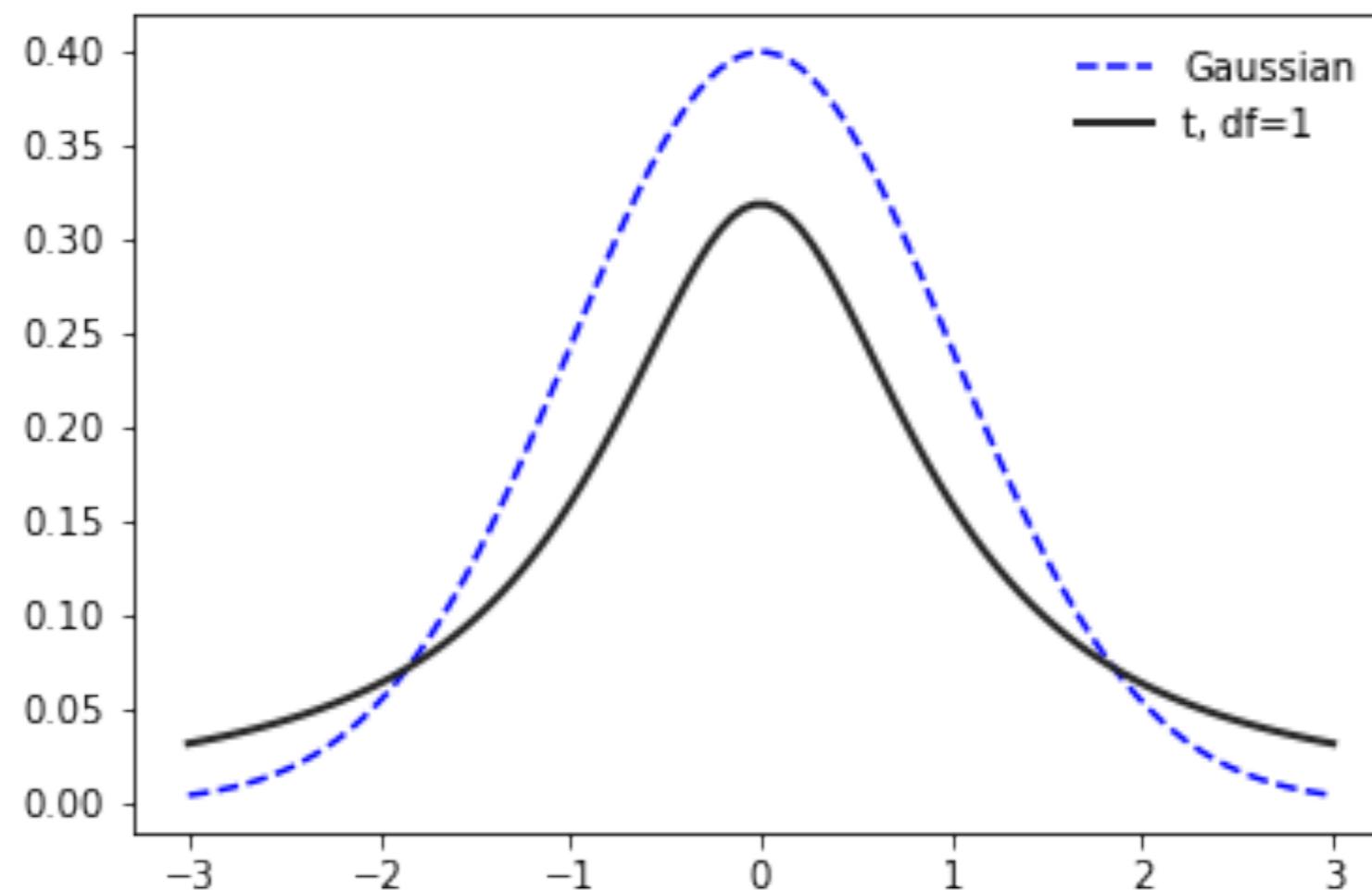
# t-SNE preserves local similarity

- $x_1, \dots, x_n$  in high dimensional space
- Map to points  $y_1, \dots, y_n$  in low dimensional space
- Gaussian centred at  $x_i$  gives same points as Student t-distribution centred at  $y_i$
- 



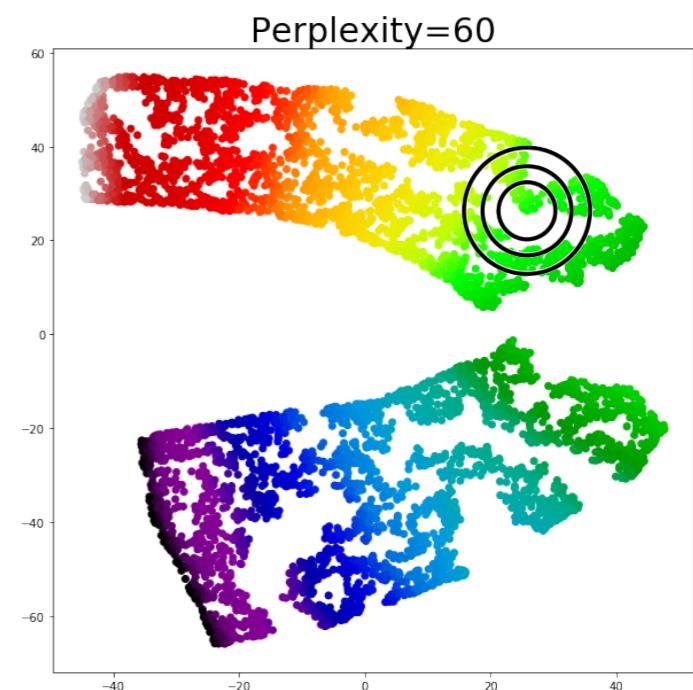
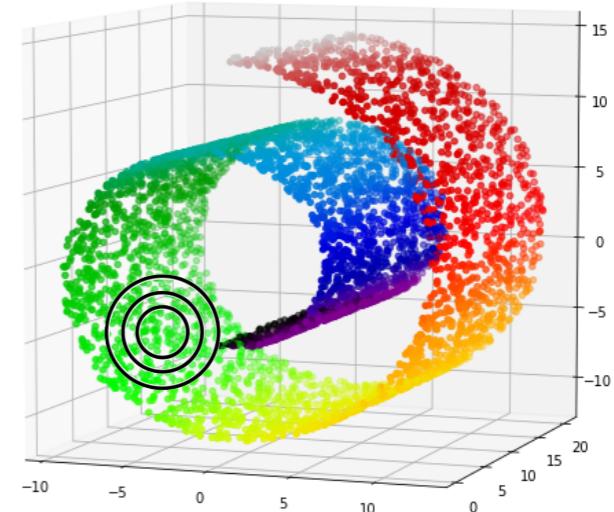
# Why Student t-distribution for lower-dimensional points?

- Faster to evaluate than Gaussian
- More weight on tails
  - Spreads out low dimensional points



# t-SNE algorithm

1. Calculate probability of selecting pair of points  $x_i, x_j$
2. Initialise  $y_1, \dots, y_n$
3. Calculate probability of selecting pair of points  $y_i, y_j$
4. Move  $y_1, \dots, y_n$  to decrease distance between probability distributions
5. Repeat 3-4 until convergence



# t-SNE algorithm

1. Calculate probability of selecting pair of points  $x_i, x_j$
2. Initialise  $y_1, \dots, y_n$
3. Calculate probability of selecting pair of points  $y_i, y_j$
4. Move  $y_1, \dots, y_n$  to decrease distance between probability distributions
5. Repeat 3-4 until convergence

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$$
$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma_i^2\right)}{\sum_{l \neq i} \exp\left(-\|x_i - x_l\|^2/2\sigma_i^2\right)}$$

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{l \neq i} \left(1 + \|y_i - y_l\|^2\right)^{-1}}$$

$$KL(P||Q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

# Hyperparameters

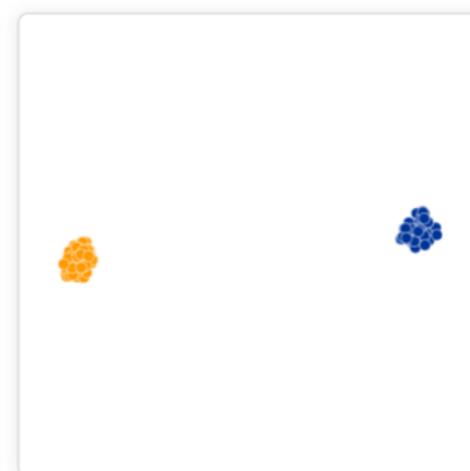
- Parameters for gradient descent
- Perplexity = smooth measure of number of neighbours
  - Number of neighbours roughly equal for each  $x_i$
  - Sigma differs for each  $x_i$



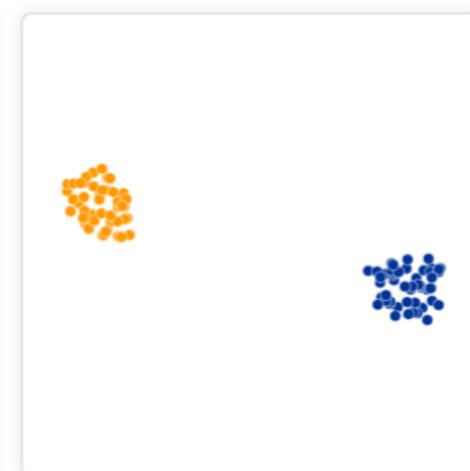
Perplexity: 2  
Step: 5,000



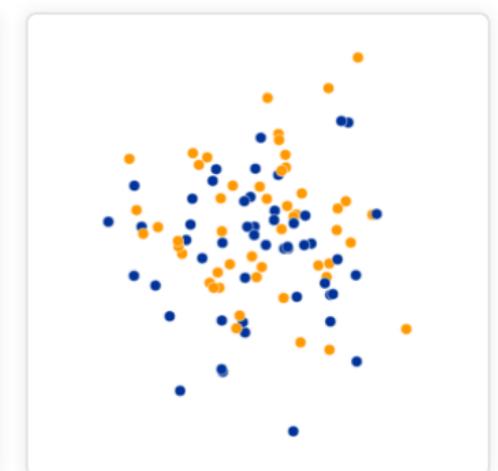
Perplexity: 5  
Step: 5,000



Perplexity: 30  
Step: 5,000



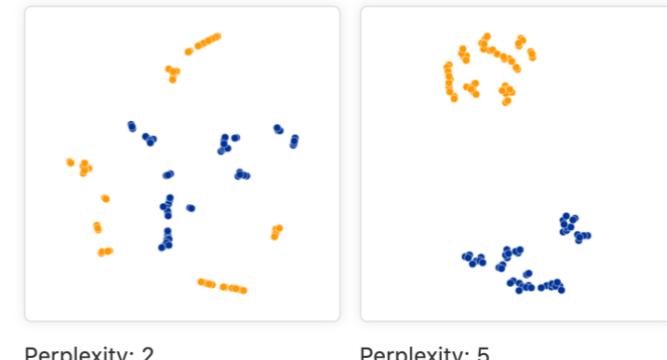
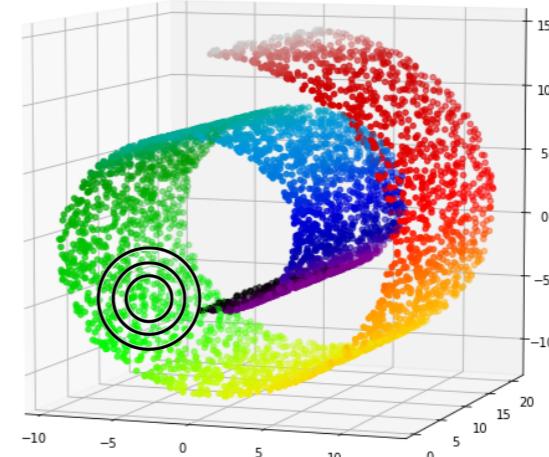
Perplexity: 50  
Step: 5,000



Perplexity: 100  
Step: 5,000

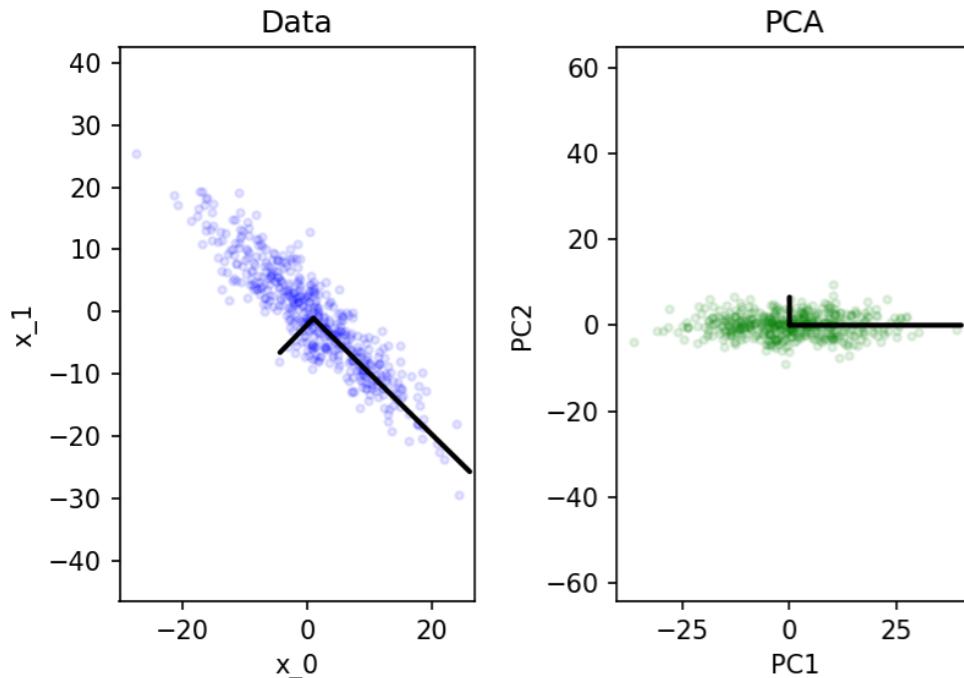
# t-distributed Stochastic Neighbour Embedding

- Preserves local similarity
- Stochastic
- Hyperparameters
  - Perplexity
- No explicit mapping

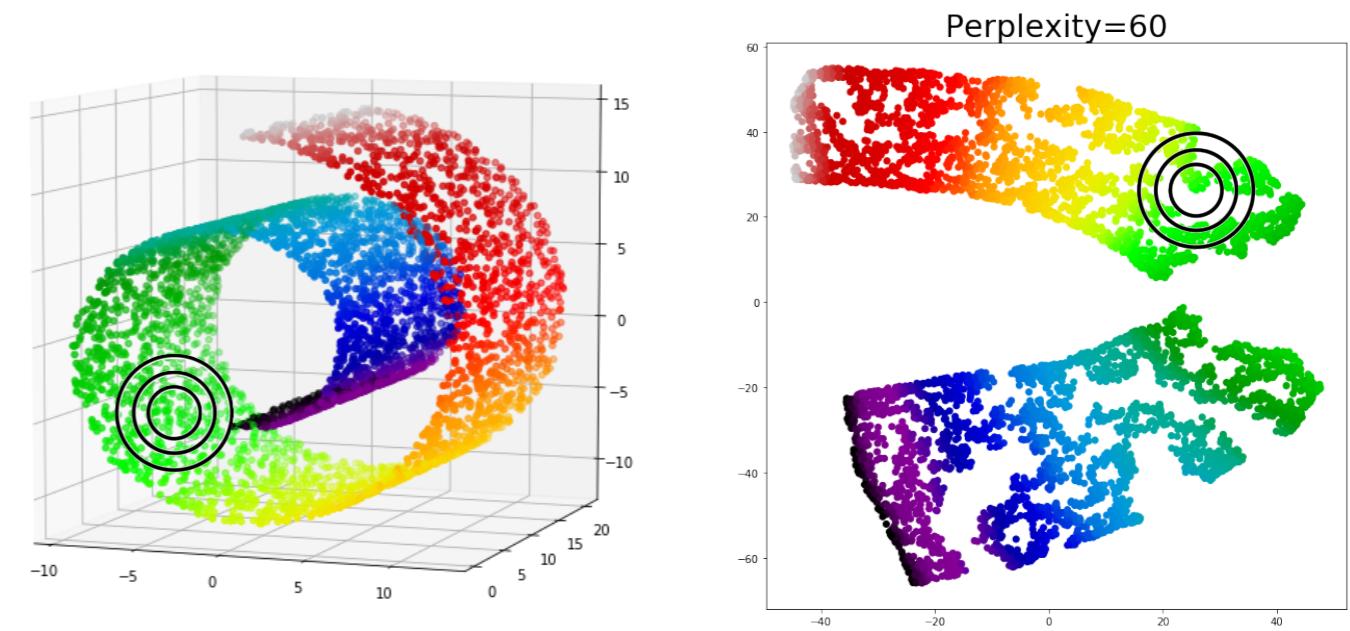


# Dimensionality reduction

## Principal Components Analysis



## t-Stochastic Neighbour Embedding



## Sparse Factor Analysis

$$Y = \Lambda X + \varepsilon$$

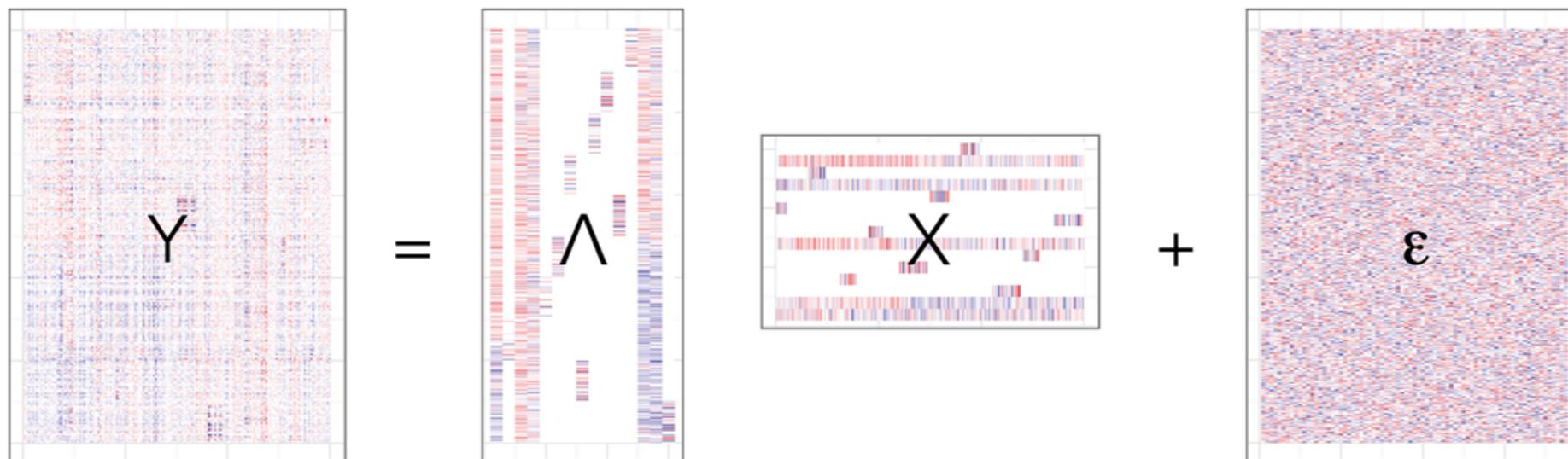
A diagram illustrating Sparse Factor Analysis. On the left is a dense matrix labeled  $Y$ . An equals sign follows, then a sparse matrix labeled  $\Lambda$ , then a sparse matrix labeled  $X$ , then a plus sign, and finally a sparse matrix labeled  $\varepsilon$ .

$$\begin{matrix} \text{blue} & \text{white} \\ \text{white} & \text{blue} \end{matrix} = \begin{matrix} \text{blue} & \text{white} \\ \text{white} & \text{white} \end{matrix} + \begin{matrix} \text{white} & \text{blue} \\ \text{white} & \text{white} \end{matrix} + \begin{matrix} \text{white} & \text{white} \\ \text{blue} & \text{white} \end{matrix}$$

Kath Nicholls

# Sparse Factor Analysis

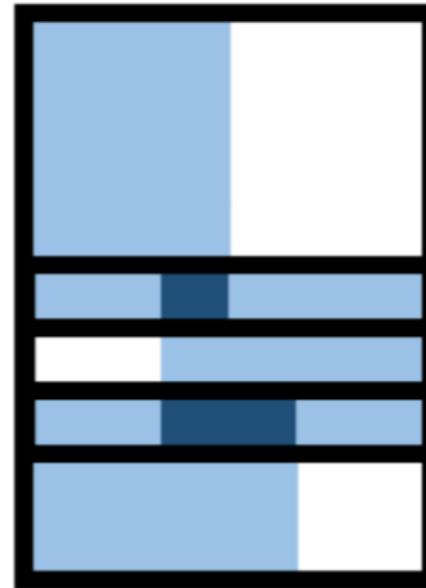
- Identify latent variables in data
  - Explicit mapping to latent space
- Stochastic
- Sparsity helps choice of k



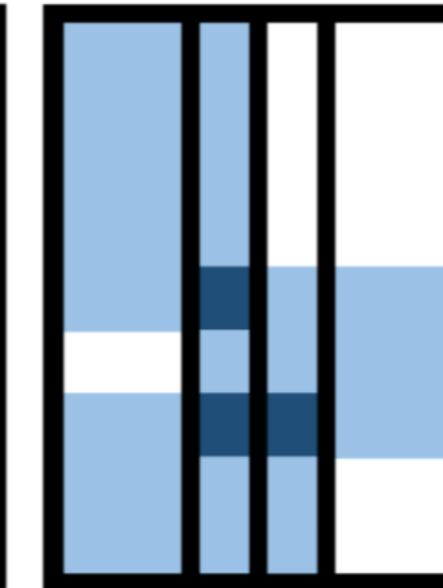
# Sparse factor analysis finds biclusters



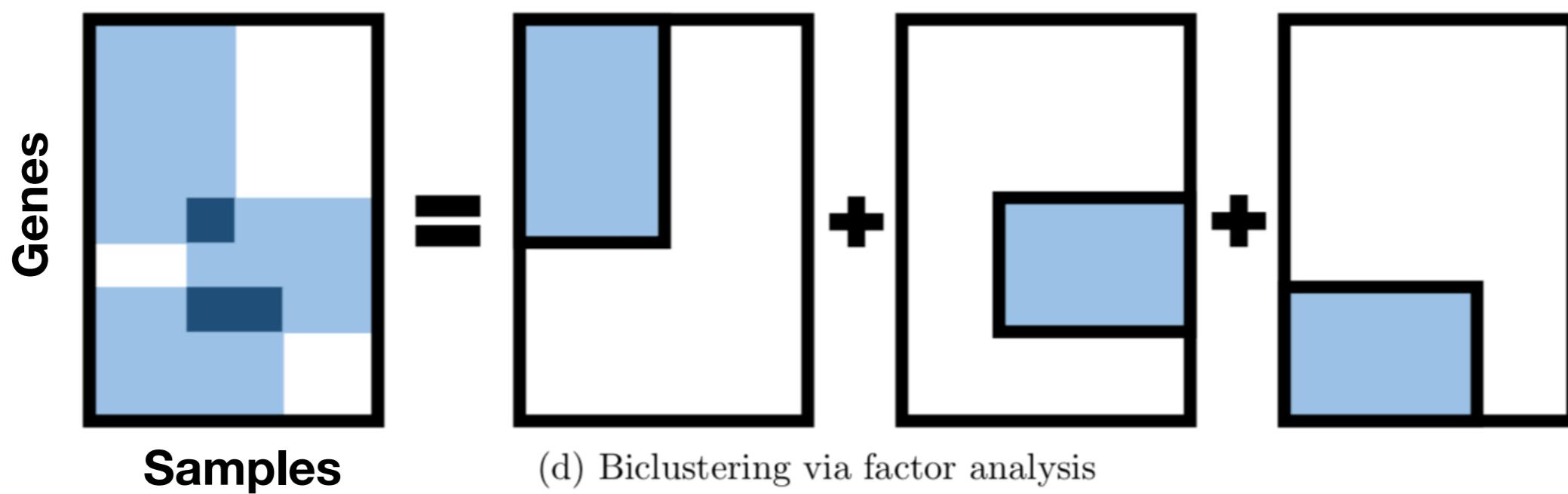
(a) Original matrix



(b) Clustering rows



(c) Clustering columns



(d) Biclustering via factor analysis

# Aim of biclustering

- Gene regulatory networks are thought to be modular
- Biclustering finds modules, and identifies links between modules and disease (and/or cell type)

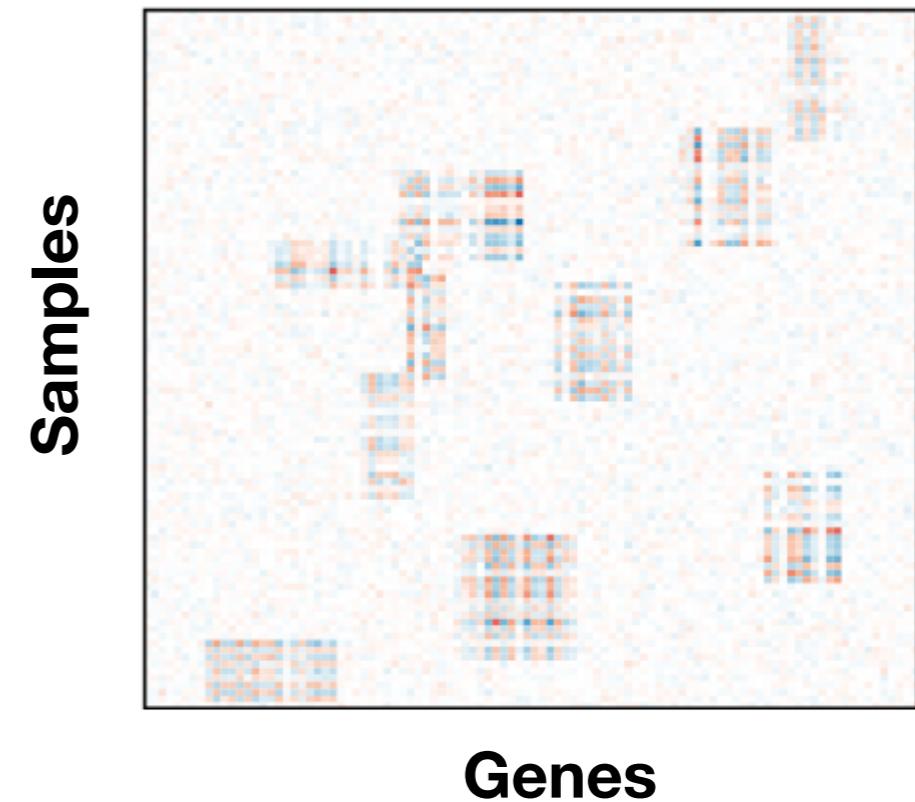
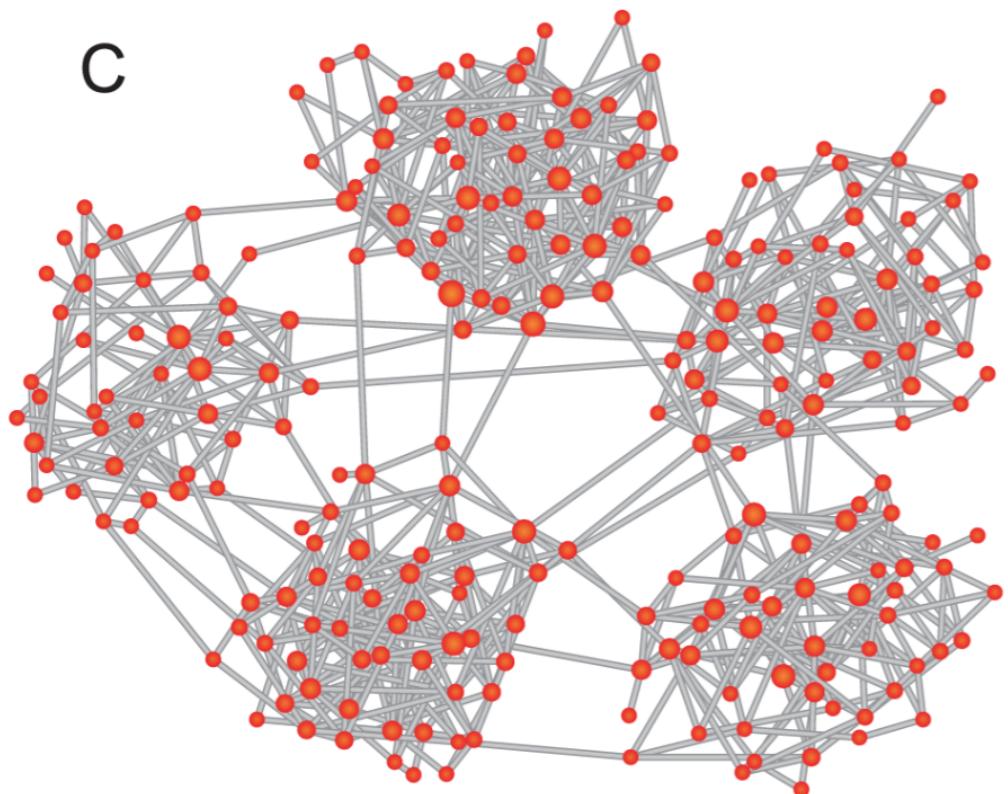


Figure 1C Müller-Linow et al. 2008; Figure 3 Moran et al. 2019

# Comparison

PCA	t-SNE	Sparse Factor Analysis
<b>Deterministic</b>	Stochastic	Stochastic
Linear	<b>Non-linear</b>	Linear
<b>Interpretable</b>	Not interpretable	<b>Interpretable</b>
<b>Mapping can be applied to other datasets</b>	Mapping only works for one dataset	<b>Mapping can be applied to other datasets</b>
<b>Easy to adjust for k</b>	Have to re-run for each k	Have to re-run for each k
Good for explicit reduction	Good for visualisation	Good for biclusters